What the 'Moonwalk' Illusion Reveals about the Perception of Relative Depth from Motion

Sarah Kromrey[1,2], Evgeniy Bart[3], Jay Hegdé[1,2,4]*

1 Brain and Behavior Discovery Institute, Georgia Health Sciences University, Augusta, Georgia, United States of America, 2 Vision Discovery Institute, Georgia Health Sciences University, Augusta, Georgia, United States of America, 3 Palo Alto Research Center (PARC), Palo Alto, California, United States of America, 4 Department of Ophthalmology, Georgia Health Sciences University, Augusta, Georgia, United States of America

Abstract

When one visual object moves behind another, the object farther from the viewer is progressively occluded and/or disoccluded by the nearer object. For nearly half a century, this dynamic occlusion cue has beenthought to be sufficient by itself for determining the relative depth of the two objects. This view is consistent with the self-evident geometric fact that the surface undergoing dynamic occlusion is always farther from the viewer than the occluding surface. Here we use a contextual manipulation ofa previously known motion illusion, which we refer to as the'Moonwalk' illusion, to demonstrate that the visual system cannot determine relative depth from dynamic occlusion alone. Indeed, in the Moonwalk illusion, human observers perceive a relative depth contrary to the dynamic occlusion cue. However, the perception of the expected relative depth is restored by contextual manipulations unrelated to dynamic occlusion. On the other hand, we show that an Ideal Observer can determine using dynamic occlusion alone in the same Moonwalk stimuli, indicating that the dynamic occlusion cue is, in principle, sufficient for determining relative depth. Our results indicate that in order to correctly perceive relative depth from dynamic occlusion, the human brain, unlike the Ideal Observer, needs additionalsegmentation information that delineate the occluder from the occluded object. Thus, neural mechanisms of object segmentation must, in addition to motion mechanisms that extract information about relative depth, play a crucial role in the perception of relative depth from motion.

Citation: Kromrey S, Bart E, Hegdé J (2011) What the 'Moonwalk' Illusion Reveals about the Perception of Relative Depth from Motion. PLoS ONE 6(6): e20951. doi:10.1371/journal.pone.0020951

Editor: Matjaz Perc, University of Maribor, Slovenia

Received April 12, 2011; **Accepted** May 13, 2011; **Published** June 22, 2011

Funding: This study was supported in part by the Brain and Behavior Discovery Institute and by the Vision Discovery Institute of the Georgia Health Sciences University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jhegde@mcg.edu

These authors contributed equally to this work.

Introduction

When one visual object moves behind another, it provides a compelling sense of their relative depth, or depth-order (*i.e.*, which object is closer to the viewer and which object is farther in depth). Depth-order from motion (DFM) is one of the strongest and the most ubiquitous cues to depth-order under natural viewing conditions[1,2,3,4,5,6,7,8,9]. Indeed, DFM can override depth-order from many other types of depth-order cues, including static occlusion[8,10]. DFM can also resolve ambiguities from other depth-order cues[8,10]. The neural mechanisms of DFM are almost entirely unknown. Thus, observations that can constrain and inform the search for the neural correlates of DFM are very valuable.

Previous studies have identified two distinct types of DFM cue (for overviews, see[8,11,12,13,14]; also see Supporting Information S1 about the biasing effects of motion shear; [15]). For instance, when a window shade is drawn shut, it progressively occludes scene elements outside the window. Conversely, when the shade is pulled open, those same scene elements are progressively disoccluded. This dynamic occlusion/disocclusion has long been recognized as a potential cue for depth-order and has been termed the accretion-deletion (AD) cue, also referred to as the dynamic- or kinetic occlusion cue[1,2,3,4,6,7,8,9]. This cue is the focus of this study. The other DFM cue, called the boundary flow cue (BF cue, also referred to as the common motion cue) results from the fact that the surface elements of the occluder move coherently with the occlusion boundary (*i.e.*, boundary between the occluder and the occluded object)[9,11,16].

For the last several decades, it has been thought that the AD cue can elicit the DFM percept by itself, *i.e.*, that AD cue is self-sufficient[1,2,3,4,6,7,8,9]. This would appear geometrically self-evident, since the surface undergoing occlusion/disocclusion is always the farther surface. The self-sufficiency of the AD cue, if valid, has important implications for the neural mechanisms by which the brain extracts DFM information from the AD cue. For instance, the underlying neural processes, and the experimental approach to finding them, will have to be fundamentally different if the brain can determine DFM solely by tracking the accretion-deletion of image elements (*i.e.*, if the AD cue is self-sufficient), *vs.* if it has to explicitly determine an additional parameter, such as the border at which accretion-deletion occurs. If depth-order cannot be determined solely from the occlusion/disocclusion of the occluded object, *e.g.*, information about the occlusion border is also needed, it will mean that DFM perception cannot be implemented solely by first-order (*i.e.*, luminance-based) motion mechanisms[1,2,3,4,6,7,9]. Moreover, it will mean that the AD cue and the BF cue are not mutually redundant. This will mean, in turn, that the current view of DFM as a process supported by two equivalent cues[1,2,3,4,6,7,9,16] is not valid. Altogether, the question of whether the AD cue is self-sufficient is crucial to understanding how we perceive depth-order from motion.

In this report, we use Ideal Observer analysis to show that the brain, in principle, could determine depth-order from dynamic occlusion alone, consistent with the longstanding belief. However, usingcontextual manipulations of a previously reported motion illusion ([17,18]; also see Demo Movies 1 and 2, downloadable from www.hegde.us/DFMdemo1.avi and www.hegde.us/DFMdemo2.avi, respectively), which we refer to as the "Moonwalk Illusion" for convenience, we empirically demonstrate that the human brain cannot determine depth-order in this fashion and that it needs additional information that segments the occluder from the occluded object.

Materials and Methods

Subjects

Eleven (6 female) adult volunteer subjects with normal or corrected-to-normal vision participated in this study. All subjects provided written informed consent prior to the study. The institutional review boardof the Georgia Health Sciences University, called the Human Assurance Committee (HAC), specifically approved this study. This investigation was conducted according to the principles expressed in the Declaration of Helsinki.

Stimuli and Task

The experiments were carried out largely as described previously (Hegdé, et al., 2004). Both the center and surround of each stimulus (Fig. 1) consisted of random dot surfaces with a dot density of 50% and, unless noted otherwise, a Michelson contrast of 1.0. The center and the surround did not physically move relative to each other. Unless noted otherwise, the surface properties of the center vs. surround were identical, so that the two surfaces were indistinguishable in any single given frame. To create a surround with a given flicker, we set the probability of a given surround dot surviving one frame to a value between 50% (random flicker) to 100% (static dots). In order to reduce the contrast of a given surface, we reduced contrast of the dots while keeping the mean luminance unchanged.

Stimuli (each 6.2O dia) were presented centered on the fixation spot on a Dell 75 Hz LCD display against a neutral gray background. The center dots all moved coherently at 6O/s in a given direction during a given trial, but the direction of motion varied randomly from trial to trial. Subjects viewed the stimuli for 4 s while maintaining fixation (with blinks as necessary), and reported the depth-order of the center using a key press. Subjects were told to report the depth-order they perceived without regard to what the expected or 'correct' depth-order was. No feedback was provided. All trials were randomly interleaved.

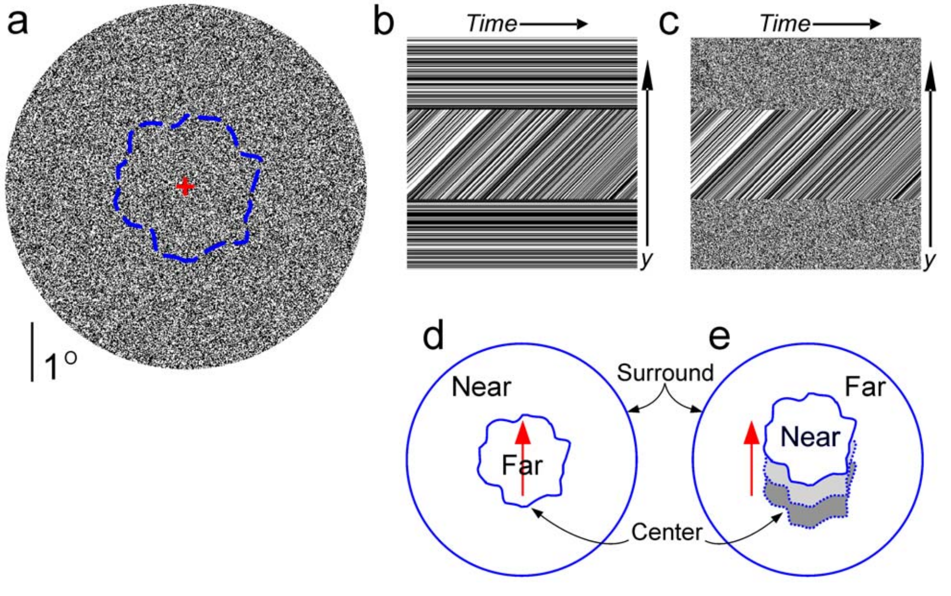


Figure 1. Stimuli and percepts. (a) A single frame of a typical motion stimulus. Unless noted otherwise, each stimulus consisted of two random dot surfaces: a center surface with an irregular outline (imaginary blue line) surrounded by an annulus. Also, the annulus, as well as the outline of the center, remained stationary in all stimuli. Subjects reported the depth-order of the center using a key press. (b,c) Space-time (ST) plots [5,9] of the two main motion stimuli used in this study (see Supporting Information S1 for the details of ST plot construction). Depending on the condition, the dots of the surround (top and bottom strips) remained static (b) or flickered to various degrees (stimulus in (d) denotes maximal flicker). In all stimuli, the dots of the center translated smoothly (upward in the case shown), and this motion was pixel-to-pixel identical across all stimuli regardless of the surround flicker, as denoted by the fact that the lines in the middle strip are identical between (b) and (c). Also see Demo Movies 2 and 1, downloadable from www.hegde.us/DFMdemo2.avi and www.hegde.us/DFMdemo1.avi, respectively. (d,e) Percepts elicited by the stimuli in (b) and (c), respectively. When the surround dots were static, the center was perceived as a moving far surface visible through a hole in the nearer surround (d). When the surround dots were flickering, the center appeared to be the near surface moving over the farther surround (e). The effect in (e) was previously reported by Anstis and Ramachandran [17,18]. The effect in (d) is original to this study to our knowledge. Note that the depth-order reversal of the center is entirely a contextual effect, in that it occurs solely as a result of the changes in the flicker of the surround in the total absence of changes in the center.
doi:10.1371/journal.pone.0020951.g001

Estimating optic flow

Optic flow was estimated using the algorithm of Horn et al (Horn &Schunck, 1981) using software custom-written in Matlab (Mathworks Inc., Natick, MA).

Results

Depth-order Perception in Moonwalk Stimuli

The stimulus consists of two random dot surfaces: A central surface of coherently moving random dots surrounded by an annulus of flickering random dots. Although both the surfaces themselves are stationary, *i.e.*, the outline of neither surface actually moves, the central surface (or 'figure') appears to translate in the direction of the moving random dots ([17,18]; see Demo Movie 1, downloadable from www.hegde.us/DFMdemo1.avi). This motion illusion, which we refer to as the Moonwalk illusion, also has a depth-order dimension: The central figure appears to be nearer to the viewer than the flickering surround (see below). This DFM percept is the opposite of that expected from the AD cue arising from the dots of the figure undergoing accretion/deletion at the border between the center and the surround. Therefore, this illusion offers an excellent test case for studying the factors that influence DFM by the AD cue.

Our stimuli consisted of various contextual manipulations of the Moonwalk stimuli that left the accretion/deletion of the center dots completely unaffected. Except where noted otherwise, the stimuli consisted of an irregular central disc of moving random dots surrounded by stationary annulus of similar random dots (Fig. 1).

To ascertain that the AD cue in our stimuli was indeed capable of eliciting the DFM percept predicted by the AD cue, we first tested a version of this stimulus in which the surround dots were static (Fig. 1a; also see Demo Movie 2, downloadable from www.hegde.us/DFMdemo2.avi). In this stimulus, the AD cue is the sole depth-order cue, caused by the occlusion-disocclusion of the dots of the moving center by the stationary occluder *i.e.*, the surround. Specifically, the BF cue is absent in this stimulus, since this cue arises only when the occluder itself moves[8,9,11,16], whereas the occluder is stationary in this case. Since the texture elements undergoing accretion-deletion belong to the center, the predicted depth-order is that the center is perceived as the far (occluded) surface, and the surround is perceived as the near (occluding) surface. When the surround dots are static (*i.e.*, 100% coherent from one frame to the next), this is indeed the reported percept (binomial proportions test, $p<<0.01$; Fig. 2a; also see Supporting Information S1).

We then made this stimulus progressively closer to the original Moonwalk stimulus by introducing flicker to the surround while leaving the center unchanged (Fig. 1b; also see Demo Movie 1, downloadable from www.hegde.us/DFMdemo1.avi). We hypothesized that if the AD cue is sufficient by itself for DFM perception, *i.e.*, if the visual system can determine the depth-order solely by measuring the accretion/deletion of the pixels of the center, then manipulations of the surround that leave the center entirely unaffected should leave the DFM percept unaffected. Note that the center in this stimulus was pixel-to-pixel, frame-to-frame identical to the stimulus shown in Fig. 1a, so that the available accretion-deletion information was identical between the two stimuli. If the AD cue were self-sufficient for determining the DFM percept, this stimulusis expected to elicit the same depth-order percept as the stimulus with the static surround.

However, with the flickering surround, the DFM percept reversed, in that the subjects perceived the center as the nearer surface. Moreover, the variations in the flicker accounted for all non-random variation in the DFM percept (logistic regression, $r2 = 0.83$; $p<0.05$; chi-square test for the normality of the residuals, $p>0.05$). Together, these results indicate not only that the AD cue was not sufficient by itself to account for the observed DFM percept, but the AD cue along with the 'gating' information in the surround were sufficient.

AD Cue is Self-Sufficient from the Computational Viewpoint

One potential concern about interpreting the above results as evidence of perceptual insufficiency of the AD cue is that it may not be possible to determine depth-order solely by measuring the accretion and/or deletion of the occluded object to begin with. This would mean that thedefinition of the AD cue as solely a function of the accretion/deletion of the occluded object without reference to the occluder *per se*, although widely accepted in the field [1,2,3,4,5,6,7,8,9], may nonetheless be artificially narrow.

To verify that the information from the accretion/deletion of the occluded object is, from a computational point of view, self-sufficient for determining DFM unambiguously, we carried out an Ideal Observer analysis (see Supporting Information S1

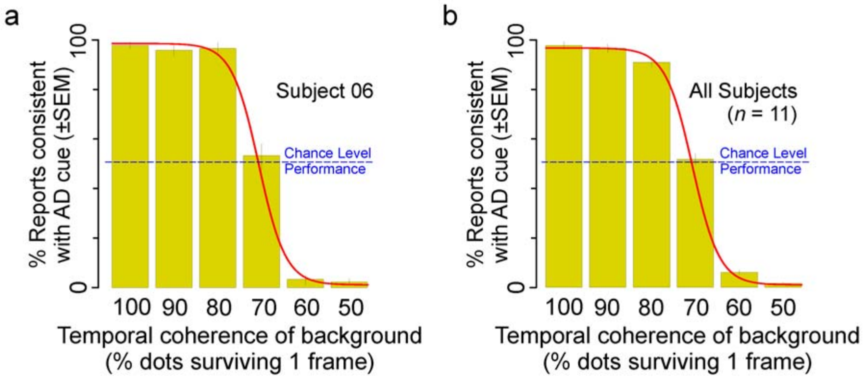


Figure 2. Dependence of the AD cue on contextual information. The temporal coherence (i.e., amount of flicker) of the surround was systematically modulated while the amount of accretion-deletion information was held constant, as described in Materials and Methods. Increasing values of temporal coherence denote decreasing flicker, with 100% temporal coherence denoting static dots. The reported depth-order percepts are shown as the percentage of trials in which the depth-order percept was consistent with the percept predicted by the AD cue by itself, i.e., that the center surface was far. (a) Reported percepts of a representative subject, and (b) average across all subjects. The red line in either panel denotes the best-fitting logistic regression line.
doi:10.1371/journal.pone.0020951.g002

for details). We found that the Ideal Observer can indeed determine DFM solely using the accretion/deletion of the center, regardless of the surround. Briefly, the Ideal Observer need only to evaluate the pixels i in the image region where a portion of the pixels get deleted over two given successive frames I^1 and I^2 (which we refer to as area 3; see Figure S2 in Supporting Information S1) to determine the log-likelihood ratio

$$L(p_c) = \log \frac{p\left(I^2_{[area\ 3]}\,|I^1,F\right)}{p\left(I^2_{[area\ 3]}\,|I^1,N\right)},$$

where F and $\mathcal{N}$ are the depth-order models where the center is far or near, respectively, and p_c is the parameters of the Ideal Observer model. Namely, p_c is the probability with which a given center pixel switches (from on to off, or vice versa) from frame 1 to frame 2. The Ideal Observer analysis also shows that the strength of the DFM information is not affected by flicker (see eq. 12 in Supporting Information S1). Thus, purely from an information processing viewpoint, it is possible to extract DFM information from the accretion-deletion information alone.

We also independently verified, using conventional optic flow algorithms[19] to analyze the two types of motion stimuli shown in Fig. 1, that the changes in the flicker did not affect the optic flow information in the center (Fig. 3a *vs.* b). The sub-region of the center in which the dots underwent accretion/deletion was also readily identifiable regardless of the surround flicker (Fig. 3c *vs.* d; see legend for additional details). Together with the Ideal Observer analysis, these results demonstrate that the available AD information is, in principle, sufficient to support the determination of DFM, even though the visual system is unable to exploit this information to determine DFM. These computational analyses also suggest that our psychophysical results are not a semantic side effect of defining the AD cue too narrowly, *i.e.*, separately from the boundary information.

'Gating' of the AD Cue by Segmentation Information

An inspection of the motion stimulus represented in Fig. 1c (see Demo Movie 1, downloadable from www.hegde.us/DFMdemo1.avi) indicates that when the surround is flickering, it becomes perpetually difficult to delineate the border between the center and surround. This suggests that one reason why the AD cue is ineffective with the flickering surround is that the visual system, unlike the Ideal Observer, needs a mechanism for delineating the occluder in order to make use of the AD cue. In other words, although the visual system cannot determine the depth-order using the AD cue alone, it can do so when the AD cue is augmented by information about the occluder. If this is true, then the DFM percept expected from the AD cue should be restored, notwithstanding the surround flicker, when center-surround segmentation is made easier by an extraneous segmentation cue.

To test this hypothesis, we carried out an additional experiment, where we made the surround more readily distinguishable from the center by changing the luminance contrast of the dots either within the center or within the surround, while leaving the other surface unchanged (Fig. 4, inset). We then measured the perceived

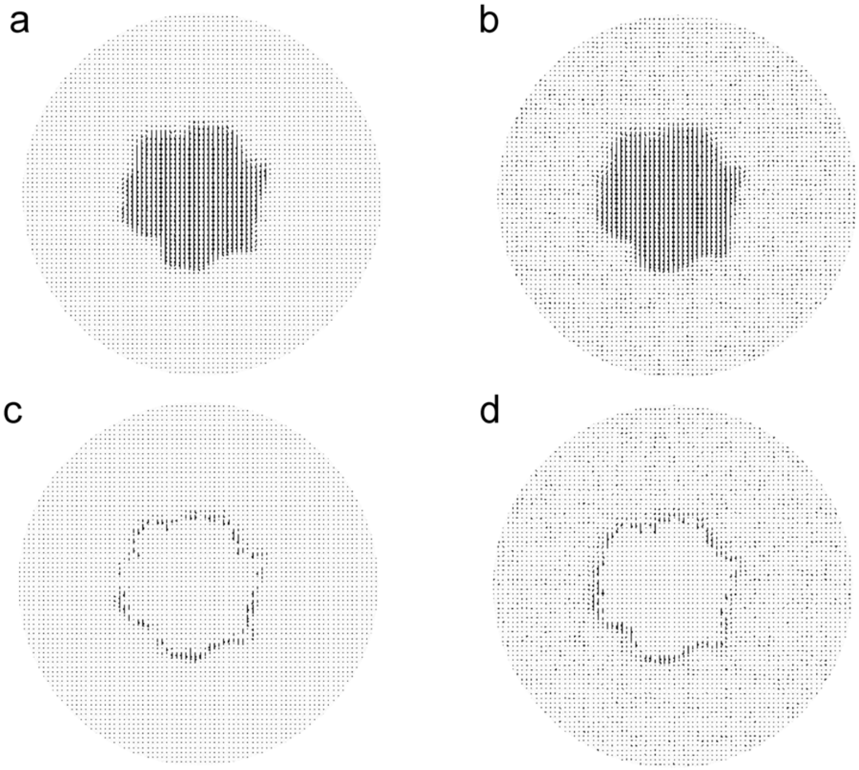


Figure 3. Optic flow information in the center with or without surround flicker. Motion information of our stimuli was estimated using a conventional optical flow estimation algorithm [19]. (a,b) Optical flow field when the surround was static (a) or flickery (b). (c,d) Accretion-deletion zone, or the region of the center where the dots underwent accretion/deletion from one frame to the next, estimated with a static surround (c) or flickery surround (d). Note that the estimated surround flicker has little effect on the estimated optic flow or the estimated accretion-deletion zone of the center itself even with this relatively simple optic flow algorithm, although corresponding estimates of the surround are somewhat different between the two conditions.
doi:10.1371/journal.pone.0020951.g003

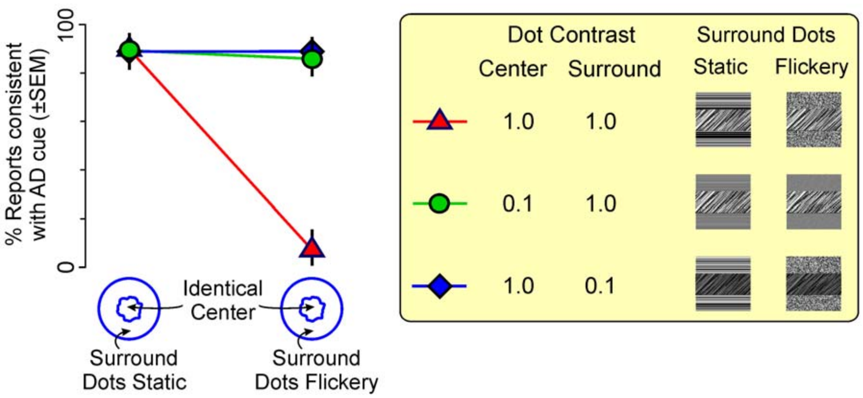


Figure 4. Restoration of the AD percept by the addition of center-surround segmentation cues. Stimuli shown in Fig. 1 were re-tested with or without additional segmentation cues to enhance the delineation of the center vs. surround. Three pairs of conditions were used (inset): with the dots in the surround at a lower contrast (blue diamonds), with the dots in the center at a lower contrast (green circles), or no contrast manipulations (red triangles; condition identical to that in Fig. 1, used as a control). The depth-order percept elicited by each of the conditions is shown as a proportion of reports consistent with the AD cue, i.e., that the center is farther than the surround. Note that the contrast manipulations do not add depth-order or motion information, because the difference in contrast is not a depth-order cue or a motion cue. The surround dots were either static or flickery (100% or 50% temporal coherence, respectively). The flicker of the surround dots was unaffected by the contrast manipulations.

doi:10.1371/journal.pone.0020951.g004

depth-order as a function of the flicker of the surround. When the surround dots were static (Fig. 4, left column), this manipulation made no difference; the perceived depth-order was consistent with the AD cue, as in the earlier experiment. When the surround dots were flickery and the dot contrast was the same between the center and the surround (red triangle, right column), the center was perceived as near, also as expected from the earlier experiment. However, when the center-surround segmentation was made easier by lowering the contrast of surround dots while leaving the center unchanged, the depth-order expected from the AD cue was restored, even though the surround flicker was unchanged (green circle at right; binomial proportions test, $p<0.05$). This restoration was not a function of the lower dot contrast in the surround *per se*, because the same restoration occurred when the center-surround distinction was mediated by the lower contrast in the center, instead of in the surround (blue diamond at right; binomial proportions test, $p<0.05$). Note that this restoration of the DFM percept predicted by the AD cue is not attributable to the contrast manipulations *per se*, since this provides no depth-order information whatsoever. Moreover, the same effect was obtained when color or luminance, instead of contrast, was used as the segmentation cue (data not shown).

The fact that the effect of the AD cue can be 'gated' by independently manipulating the border delineation indicates that the AD cue is indeed distinct from the border delineation. The fact that the predicted AD percept can be restored by solely better delineating the occluder without changing its flicker indicates that the flicker itself does not affect the accretion-deletion cue. It also indicates the failure of the AD cue in stimuli with flickery surrounds is not due to trivial causes, such as the inability to resolve individual pixels. Together, these findings support the aforementioned computational results, and show that the visual system needs additional information about the occlusion border in order to use the occlusion/disocclusion information.

Discussion

Our results demonstrate that it is possible, in principle, to unambiguously determine the depth-order of moving objects solely by keeping track of the accretion/deletion of the occluded object. However, we also show empirically that the human brain is unable to do determine depth-order in this fashion, and that the accretion-deletion information by itself is ambiguous to the visual system. The visual system needs additional constraining information about the occlusion border.

Moonwalk Illusion as Bayesian 'Explaining Away'

The percepts elicited by our stimuli, including the depth-order effects, can be readily explained as a well-known type of Bayesian inference called 'explaining away'[20,21]; also see [12,14]. Briefly, explaining away refers to a scenario where the stimulus supports two alternative interpretations, either one of which is plausible in the absence of additional constraining evidence. But when the constraining evidence is available, one of the two original interpretations becomes much more plausible. In the present case, the two plausible interpretations of our stimuli are shown in Fig. 1d and 1e, respectively. The AD cue is consistent with only one of the interpretations (Fig. 1d), but our results show that the visual system cannot use this information by itself. The constraining evidence is provided by the segmentation cue, and this additional evidence makes the interpretation in Fig. 1d much more plausible. Thus, when strong enough segmentation information is available, the brain favors the interpretation consistent with the AD cue, using the segmentation information to explain away ambiguities in the incoming AD information.

In the absence of strong enough segmentation cues, the brain is unable to use the AD cue by itself to determine DFM. Hence it chooses the alternative information where the center surface is perceived as translating in the near plane (Fig. 1e). Note that, in the absence of usable evidence that the center is occluded, translation of the center accounts for the dot motion in the center. The perception of the center as near in this case is also aided by the built-in perceptual bias to interpret the shearing created by translating surfaces as nearness cue[15]. In this case, in the absence of strong enough gating information, *i.e.*, segmentation cues, the brain chooses an interpretation to account for the remaining available information.

Implications for the Neural Substrates of DFM

Our results also provide useful constraints on the neural mechanisms by which the brain processes the AD cue. The fact that unlike the Ideal Observer, the visual system cannot use the AD cue by itself to determine DFM in a "bottom-up" fashion, suggests that the extraction of AD information is closely associated with segmentation processes. To the extent that this involves a comparison of the relative velocities of the image elements undergoing accretion/deletion *vs.* the border between the two surfaces[22,23,24], the underlying process is, by definition, a second-order motion process[9,25,26,27,28,29]. In other words, the AD cue cannot be processed solely by first-order (*i.e.*, luminance-based) motion mechanisms; second-order (*i.e.*, non-luminance based) mechanisms must be involved. Our results also predict that the neural mechanisms of AD cue processing will be different when the occluder is stationary *vs.* moving, even when the accretion-deletion information itself is identical between the two conditions. This is because occluder motion typically gives rise to the BF cue, and the motion of the occlusion boundary creates strong segmentation cues.

Our results disprove the conventional view, nearly half a century old, that the AD cue is self-sufficient for the perception of DFM[1,2,3,4,6,7,8,9]. In doing so, they offer a new perspective of the relative roles of the DFM cues in DFM perception. Contrary to conventional wisdom, the two known DFM cues are not functionally equivalent (*i.e.*, not mutually redundant), but instead play different, perhaps complementary, roles in the perception of depth-order from motion *vs.* surface segregation. As noted above, the BF cue pertains to the motion of the occluder, and the AD cue pertains to the motion of the occluded object. The two cues also engage the first *vs.* second-order motion systems differently[1,2,3,4,6,7,8,9]. Since the two cues tend to co-occur under natural viewing conditions[1,2,3,4,6,7,8,9], one cue may serve to compensate for the ambiguities in the other.

Supporting Information

Supporting Information S1 Contains various inter-related lines of evidence, including the Ideal Observer analysis, that support the findings presented in the main text.
(DOCX)

Acknowledgments

We gratefully acknowledge the help of Mr. Joel Covar, Ms. Ann Gambill and Ms. Shalon Howard during various phases of this study. We thank Drs. Christos Constantinidis and Lawrence Snyder for helpful suggestions.

Author Contributions

Conceived and designed the experiments: SK EB JH. Performed the experiments: SK. Analyzed the data: EB JH. Contributed reagents/materials/analysis tools: EB JH. Wrote the paper: SK EB JH. Designed the software used in analysis: EB JH.

References

1. Michotte A, Thines G, Crabbe G (1964) Les complements amodaux des structures perceptives. Studia psychologia: Louvain: Publications Universitaires de Louvain.
2. Kaplan GA (1969) Kinetic disruption of optical texture: the perception of depth at an edge. Perception and Psychophysics 6: 193–198.
3. Gibson JJ, Kaplan GA, Reynolds HEN, Wheeler K (1969) The change from visible to invisible: A study of optical transitions. Percept Psychophys 5: 113–116.
4. Thompson WB, Mutch KM, Berzins VA (1985) Dynamic Occlusion Analysis in Optical Flow Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 7: 374–383.
5. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2: 284–299.
6. Mutch KM, Thompson WB (1988) Analysis of accretion and deletion at boundaries in dynamic scenes. In: Richards W, ed. Natural Computation. Cambridge, MA: MIT Press. pp 44–54.
7. Niyogi SA (1995) Detecting kinetic occlusion. Massachusetts Institute of Technology, Cambridge, Massachusetts. pp 1044–1049.
8. Howard IP, Rogers BJ (2002) Seeing in Depth. Vol. 2. Depth Perception: I. Porteous, Toronto.
9. Hegdé J, Albright TD, Stoner GR (2004) Second-order motion conveys depth-order information. J Vis 4: 838–842.
10. Ono H, Rogers BJ, Ohmi M, Ono ME (1988) Dynamic occlusion and motion parallax in depth perception. Perception 17: 255–266.
11. Yonas A, Craton LG, Thompson WB (1987) Relative motion: kinetic information for the order of depth at an edge. Percept Psychophys 41: 53–59.
12. Fleet DJ, Black MJ, Nestares O (2002) Bayesian Inference of Visual Motion Boundaries. In: Lakemeyer G, Nebel B, eds. Exploring Artificial Intelligence in the New Millenium. San Francisco, CA.: Morgan Kaufmann. pp 139–173.
13. Fleet DJ, Langley K (1994) Computational analysis of non-Fourier motion. Vision Res 34: 3057–3079.
14. Fleet DJ, Weiss Y (2005) Optical flow estimation. In: Paragios N, Chen Y, Faugeras O, eds. Mathematical models for Computer Vision: The Handbook: Springer. pp 1–24.
15. Royden CS, Baker JF, Allman J (1988) Perceptions of depth elicited by occluded and shearing motions of random dots. Perception 17: 289–296.
16. Craton LG, Yonas A (1990) Kinetic occlusion: further studies of the boundary-flow cue. Percept Psychophys 47: 169–179.
17. Ramachandran VS, Anstis SM (1990) Illusory displacement of equiluminous kinetic edges. Perception 19: 611–616.
18. Anstis SM (1989) Kinetic edges become displaced, segregated, or invisible. In: Lam DM-K, Gilbert CD, eds. Proceedings of the Second Retina Research Foundation Conference, Portfolio Press, Texas. pp 247–260.
19. Horn BKP, Schunck BG (1981) Determining optical flow. Artificial Intelligence 17: 185–203.
20. Pearl J (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA: Morgan Kaufmann.
21. Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol 55: 271–304.
22. Andersen GJ, Cortese JM (1989) 2-D contour perception resulting from kinetic occlusion. Percept Psychophys 46: 49–55.
23. Beck C, Ognibeni T, Neumann H (2008) Object segmentation from motion discontinuities and temporal occlusions–a biologically inspired model. PLoS ONE 3: e3807.
24. Bruno N, Bertamini M (1990) Identifying contours from occlusion events. Percept Psychophys 48: 331–342.
25. Chubb C, Sperling G (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. J Opt Soc Am A 5: 1986–2007.
26. Cavanagh P, Mather G (1989) Motion: the long and short of it. Spat Vis 4: 103–129.
27. Albright TD (1992) Form-cue invariant motion processing in primate visual cortex. Science 255: 1141–1143.
28. Baker CL, Jr, Mareschal I (2001) Processing of second-order stimuli in the visual cortex. Prog Brain Res 134: 171–191.
29. Vaina LM, Soloviev S (2004) First-order and second-order motion: neurological evidence for neuroanatomically distinct systems. Prog Brain Res 144: 197–212.

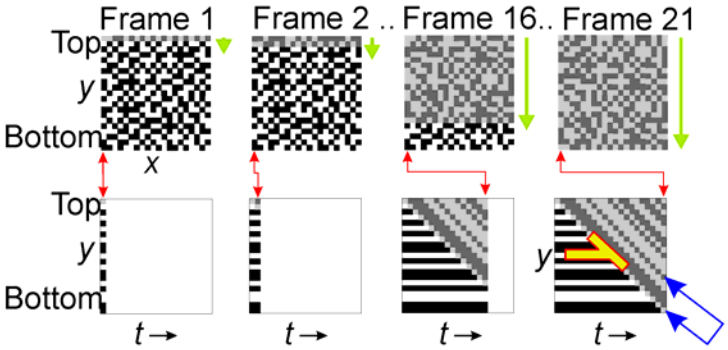
What the 'Moonwalk' Illusion Reveals about the Perception of Relative Depth from Motion

Sarah Kromrey, Evgeniy Bart, and Jay Hegdé

SUPPORTING INFORMATION

―――

Section S1. Using space-time (ST) plots as static 2-D representations of motion stimuli



Supplemental Figure S1. Space-time (ST) plots of motion stimuli.

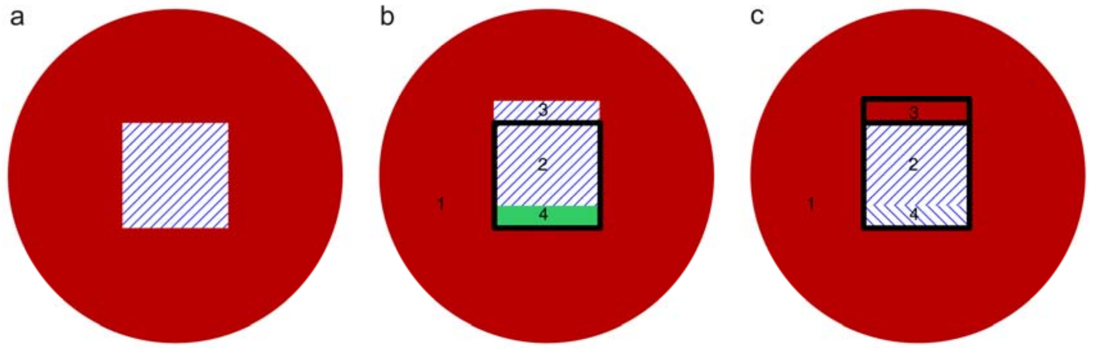
ST plots are often used to represent motion stimuli in a static 2-D form [1,2]. Figure S1 illustrates the construction of ST plots using an exemplar random dot movie with 21 frames. The top row shows four selected frames of the movie; the bottom row shows the corresponding stages of the ST plot construction. The icon at bottom right denotes the finished ST plot.

The ST plot is constructed by concatenating a designated column of pixels from successive frames of the motion stimulus, so that successive columns of the ST plot represent the temporal progression of a given set of pixels in the motion stimulus. For the stimulus shown, the top panel (highlighted in gray) slides downward over the stationary bottom panel in successive frames (green downward arrows). In the present case, the ST plot represents the temporal progression of the far left column of pixels of the movie. However, the choice of which column (or row) to use depends on the user.

When the motion stimulus contains AD and/or BF cues (see main text), the ST plot will show the corresponding diagnostic features. In the present case, the horizontal and oblique lines in the ST plot correspond to stationary and moving textures, respectively. The accretion-deletion (AD) cue is indicated by the termination of the horizontal lines (highlighted by the yellow T) at the boundary (*i.e.*, where the two sets of lines meet) between the two panels [2]. The BF cue is reflected by the fact that the orientation of the boundary (which denotes the velocity of the boundary) and that of the surface shown in dark gray (which denotes the velocity of this surface) are exactly parallel (paired arrows; also see ref. [2]). Note that the BF cue is a property of, and denotes, the near surface, whereas the AD cue denotes the far surface.

Section S2. Ideal Observer Analysis of DFM

S2.1. Notation



Supplemental Figure S2: Image notation. (**a**) Frame 1 of the motion stimulus. The center and the surround are shown as hatched and red surfaces, respectively. (**b**) Frame 2, in a scenario where the center is near (*i.e.*, in front of the surround). (**c**) Frame 2, in a scenario where the center is far (*i.e.*, behind the surround). See below for details.

The stimulus consists of a center (hatched) and a surround (red). The center is shown as rectangular in Figure S2 for clarity; however, note that the definitions and the derivation do not change if the center has arbitrary shape. Experimentally, the depth-order percepts elicited by rectangular and irregular-shaped stimuli are statistically indistinguishable (see Section S4 below).

Panel (a) shows the first frame of the sequence. Panel (b) shows the second frame of the sequence when the center is near (in front of the surround). The center has moved up. The black square is the ‘old’ footprint of the center in frame 1. The part of the center that has moved out of the old footprint is referred to as area 3. In area 3, part of the surround is occluded. The part of the center that is still inside its old footprint is area 2. The part within the old footprint which the center revealed after moving is area 4 (small green rectangle). Note that area 4 was not visible before. The remainder of the image is called area 1. Panel (c) shows the second frame of the sequence if the center is far (behind the surround). The center has moved up. The black square is the footprint of the center from frame 1. Since the center is behind, the footprint of the center remains the same in frame 2.   Area 3 (also enclosed by a black frame) is the same region in the image corresponding to area 3 in panel (b). It is in red to indicate that surround is still visible in that area. The part of the center that is still visible is area 2. This is the same image region as area 2 in panel (b), both in terms of location within the image and the visual content of the region. The new part of the center that became visible due to motion is area 4. This is the same location in the image as area 4 in panel (b). It is hatched to indicate that it belongs to the center. The hatching is in a different direction to indicate that this is a previously invisible part of the center. Everything else is area 1, and it is the same as area 1 in (b), in terms of both location within the image and the visual content. To summarize, areas 1, 2, 3, and 4 each correspond to the same locations in panels (b) and (c). So we will also use these locations when referring to frame 1 in panel (a). The visual content of these areas may be different in different images.

S2.2. Decision-making

We emphasize that the purpose of the following analyses is solely to demonstrate that it is possible, from a strictly information processing viewpoint, to extract DFM information from accretion-information alone. Since this issue is independent of whether the visual system can actually *utilize* such DFM information (if any), the following analysis does not attempt to show (or, for that matter, claim) that the underlying information processing steps are biologically feasible.

All stimuli are generated from the model where the center is far (*i.e.*, behind the surround). So if frame 1 (denoted I^1) is as in panel (a), then frame 2 (denoted I^2) is as in panel (c). The observer must decide between two models: the center is far (model denoted by F) and the center is near (model denoted by N). For this, we need to compare $p(I^1, I^2|F)$ and $p(I^1, I^2|N)$.

Now,

$$p(I^1, I^2|F) = p(I^1|F)p(I^2|I^1, F), \qquad (1)$$

and

$$p(I^1, I^2|N) = p(I^1|N)p(I^2|I^1, N). \qquad (2)$$

$p(I^1|F)$ and $p(I^1|N)$ are the prior probabilities of the first frame, computed without seeing any other images. These do not depend on the occlusion model (near or far), but only on the image model (i.e., what we think images look like). Therefore it is natural to assume $p(I^1|F) = p(I^1|N)$. Note that this is the case even when priors for the center and the surround are different, because when only one frame (the frame I^1) is considered, the amount of center pixels is the same for the two models, and similarly for the surround pixels. Under this assumption, we only need to compare $p(I^2|I^1, F)$ and $p(I^2|I^1, N)$ to decide on the F *vs.* N model. Next, we propose an ideal observer model for how this comparison is performed.

S2.3. Ideal Observer model

We assume that the center location, shape, and velocity are known. Similar assumptions are standard in comparable analyses [3]. The rationale is that these can be very accurately estimated after watching a few frames of the stimulus. One exception is the location of area 3, which may be difficult to be estimated precisely due to flicker.

We also use the 'brightness constancy' model of image appearance for the center and a flicker model for the surround. These models are detailed below.

For the center, the conventional brightness constancy assumption [4,5,6] states that for two corresponding pixels, i_1 in frame 1 and i_2 in frame 2,

$$I_{i_2}^2 - I_{i_1}^1 \sim N(0, \sigma), \qquad (3)$$

where I_i^f is the grey value of frame f at pixel i. That is, we assume that the noise is Gaussian with zero mean and standard deviation σ. Usually σ is assumed to be small. Since center does not flicker in our stimuli, this is a reasonable assumption.

Since the image is binary in our case, we modify the brightness constancy model as follows: the pixel flips with probability p_c (which is a new parameter of the model replacing σ), and remains unchanged with probability 1 – p_c . The model becomes:

$$|I_{i_2}^2 - I_{i_1}^1| \sim \begin{cases} 0 \text{ with probability } 1{-}p_c \\ 1 \text{ with probability } p_c \end{cases} \qquad (4)$$

Brightness constancy would not model the surround well. The reason is that the surround flickers, and therefore changes in appearance are expected. It is therefore desirable to replace brightness constancy with a more accurate model of surround appearance. Here, we assume a two-stage model for flicker. In the first stage, the model observes the stimulus and estimates from it the flicker probability r. In the second stage, it uses the estimated distribution of r given the data observed so far to evaluate the likelihood of the surround from subsequent frames. This appearance model monitors the number (or area) of pixels that changed between frames (biologically, this could be done using motion detectors or change detectors), and penalizes deviations of the number of pixels that actually flipped from the expected value r.

Suppose the model has observed the stimulus for several frames. Denote the total number of observed pixels by N_0 (this is the number of pixels in area 1 times the number of frames). Of these, denote the number of pixels that flipped by N_0^f. The number of pixels that didn't flip is then $N_0^c = N_0 - N_0^f$. Assuming a uniform prior on r, the posterior is $\mathrm{Beta}(r|N_0^f + 1, N_0^c + 1)$, *i.e.*, the Beta distribution with parameters $\alpha = N_0^f + 1$ and $\beta = N_0^c + 1$.

Next, the model observes area 3 for two consecutive frames. Denote the total number of pixels in area 3 by N_3. Denote the number of pixels that have flipped by N_3^f, and the number of pixels that didn't flip by $N_3^c = N_3 - N_3^f$. We assume N_3^f is distributed binomially with the probability of flip r, where r itself is distributed according to $\mathrm{Beta}(r|N_0^f + 1, N_0^c + 1)$ (the posterior estimated in the previous step). The surround model $p(N_3^f)$ is then obtained by integrating out the parameter r from the joint distribution $p(N_3^f, r)$. It is straightforward to show that

$$p(N_3^f) = \frac{\mathrm{B}(\mathrm{N}_0^\mathrm{f}+N_3^f+1, N_0^c+N_3^c+1)}{\mathrm{B}(\mathrm{N}_0^\mathrm{f}+1, N_0^c+1)} \binom{N_3}{N_3^f}. \qquad (5)$$

Note that this takes into account the uncertainty in estimating r from observations.

Again, note that the surround appearance model models the flicker rather than the image appearance directly. This is because random changes in the surround appearance are expected, and trying to predict the exact pixel-wise appearance of the subsequent frame is therefore impossible and useless.

Note that although the pixels in area 3 are independent given r, integrating r out introduces a dependency between these pixels and we no longer can model them as independent. Intuitively, the reason is that evaluating the flicker is best carried out by pooling over multiple pixels. We do, however, assume that the four image areas (Figure S2) are treated independently. The reason is that flicker can be computed from a small image patch, much larger than a single pixel, but much smaller than any of the four image areas. Such local patches can be treated independently, and due to their small size they are unlikely to span multiple image areas.

Under this assumption,

$$p(I^2|I^1, F) = p(I^2_{[\text{area 1}]}|I^1, F) \cdot p(I^2_{[\text{area 2}]}|I^1, F) \cdot p(I^2_{[\text{area 3}]}|I^1, F) \cdot p(I^2_{[\text{area 4}]}|I^1, F) \quad (6)$$

and

$$p(I^2|I^1, N) = p(I^2_{[\text{area 1}]}|I^1, N) \cdot p(I^2_{[\text{area 2}]}|I^1, N) \cdot p(I^2_{[\text{area 3}]}|I^1, N) \cdot p(I^2_{[\text{area 4}]}|I^1, N) \quad (7)$$

Now, pixels in area 1 belong to the surround under both models F and N. Therefore, $p(I^2_{[\text{area 1}]}|I^1, F) = p(I^2_{[\text{area 1}]}|I^1, N)$, and pixels in area 1 can be ignored in the comparison.

Pixels in area 2 belong to the moving center in both models F and N. Moreover, since the center moves with the same velocity in F and N, the pixels in area 2 have the same corresponding region in frame 1 for both models F and N. Therefore, $p(I^2_{[\text{area 2}]}|I^1, F) = p(I^2_{[\text{area 2}]}|I^1, N)$, and area 2 can be ignored in the comparison as well. Finally, pixels in area 4 belong to the previously invisible part of the center in model F. These pixels belong to the previously invisible part of the surround in model N. The size and location of this area is the same in models F and N (because the direction and speed are the same). So they do not depend on the motion model, but rather only on the prior information about image appearance. More specifically, in F, this is the prior on center appearance, and in N, this is the prior on surround appearance. In our case, the two priors are the same (random black-and-white pixel noise). More generally, whenever we talk about kinetic edges, the two priors must be the same—otherwise the edge between center and surround will not be purely kinetic. Therefore, $p(I^2_{[\text{area 4}]}|I^1, F) = p(I^2_{[\text{area 4}]}|I^1, N)$, and area 4 can be ignored in this comparison as well (see below).

Note that in depth model F, area 4 represents the region of the image where the center pixels are disoccluded over successive frames. Previous studies have shown that human observers can use disocclusion to determine depth-order, just as they can use occlusion [7,8,9].   It should be noted that, while our model ignores the disocclusion information (in area 4) for the sake of simplicity, it is straightforward to expand the model to incorporate this information in one or both of the following two ways. First, if the priors on center and surround appearance are different (as in the window shade example or in the experiments where the center and surround contrast was different, Fig. 4), then $p(I^2_{[\text{area 4}]}|I^1, F) \neq p(I^2_{[\text{area 4}]}|I^1, N)$, and area 4 will become informative for the decision. Second, note that the current model considers only two consecutive frames. If integrating over multiple frames is added, area 4 might again become useful for the comparison. Note also that the goal of the current analysis is to show that depth order information is still present in the stimuli, rather than to model the human decision process. Therefore modeling these additional decision factors is not necessary, since they only add, but do not reduce, the amount of information present in the stimuli. It is of interest that the subjects ignore this multi-frame information as well (as is evident from the illusion), but this is not the focus of the current paper.

In summary, to decide between F and N in the current model, we only need to compare $p(I^2_{[\text{area 3}]}|I^1, F)$ to $p(I^2_{[\text{area 3}]}|I^1, N)$, and the log-likelihood ratio is

$$L\,(p_c) \;=\; \log \frac{p(I^2_{[\text{area 3}]}|I^1,\ F)}{p(I^2_{[\text{area 3}]}|I^1,\ N)} \qquad (8)$$

Next, we compute this log-likelihood ratio for our stimuli. Note that the actual stimuli are always from the model F. Therefore, all pixels in area 3 belong to the surround. When the surround is dynamic, the pixels may flip with probability *f*. Therefore, for all pixels *i* in area 3, $I^2_i = I^1_i$ with probability 1 – f and $I^2_i \neq I^1_i$ otherwise.

We compute the probabilities of these stimuli under models F and N.

For F, the model prediction is that area 3 corresponds to the same surround region in both frames. Therefore,

$$p(I^2_{[\text{area 3}]}|I^1, F) \quad = p(N^f_3) \qquad (9)$$

$$= \frac{\mathrm{B}(\mathrm{N}^{\mathrm{f}}_0+N^f_3+1, N^c_0+N^c_3+1)}{\mathrm{B}(\mathrm{N}^{\mathrm{f}}_0+1, N^c_0+1)} \binom{N_3}{N^f_3} \qquad (10)$$

In our stimuli, the fraction of pixels that flip is *f*. Therefore, $N^f_3 = fN_3$, and $N^f_0 = fN_0$. It is reasonable to assume that $N_3 \ll N_0$. This is because N_0 consists of many pixels observed over multiple frames, whereas N_3 includes only pixels in the (relatively small) area 3 over one pair of frames. In this case, and using the Stirling approximation, we obtain

$$p\left(I^2_{[\text{area 3}]}\middle|I^1, F\right) \approx \frac{1}{\sqrt{2\pi N_3 f(1-f)}}.$$

This approximation applies to most values of f except those close to 0 (or to 1). The reason is that when $f \approx 0$, N_0 approaches 0 and the Stirling approximation becomes inaccurate.

The model N predicts that pixel i in area 3 will correspond to some pixel on the center. Denote the value of this center pixel by C_i. That is, C_i is the value of that pixel on the center which would correspond to pixel i in the second frame given the center's location and velocity. Therefore,

$$p(I^2_{[\text{area 3}]}|I^1, N) = \prod_{i \in \text{area 3}} p(I^2_i = C_i) = p_c^{N_{\text{diff}}}(1 - p_c)^{N_{\text{same}}} \qquad (11)$$

Here N_{same} is the number of pixels that agree with the prediction (*i. e.*, those for which $I^2_i = C_i$), and N_{diff} is the number of pixels that are different from the prediction. In our case, both the center and the surround consist of random dots, 50% black and 50% white. Therefore, $p(I^2_i = C_i) = 0.5$, and $N_{\text{same}} \simeq N_{\text{diff}} \simeq N_3/2$.

Therefore, the log-likelihood ratio L is

$$L(p_c) = \log p\left(I^2_{[\text{area 3}]}|I^1, F\right) - \log p_c^{\frac{N_3}{2}}(1 - p_c)^{\frac{N_3}{2}} =$$

$$-\frac{1}{2}\log(2\pi \mathrm{N}_3) - \frac{1}{2}\log f(1 - f) - \frac{N_3}{2}\log p_c(1 - p_c). \qquad (12)$$

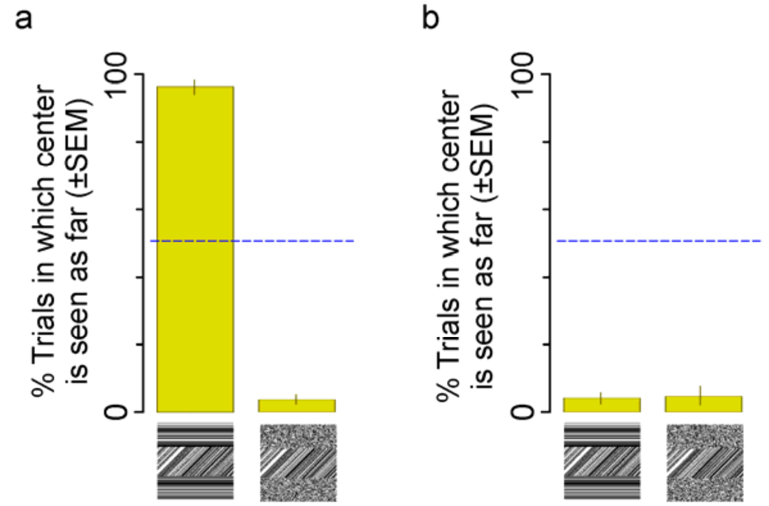
Note that this equation is dominated by the last term, which is proportional to N_3. The first term is a constant proportional to $\log N_3$. The second term is small and nearly constant. More precisely, as f changes from 0.5 to 0.01, the second term changes from 0.6 to 2.4. For comparison, if $p_c = 0.01$ and $N_3 = 100$, then the third term is equal to 231. Therefore,

$$L(p_c) \approx -\frac{N_3}{2}\log p_c(1 - p_c) = \text{const.}$$

The Ideal Observer judges the center to be far if $L > 0$, and judges the center to be near if $L < 0$.

For reasonable (*i. e.*, small) values of p_c, L is positive. Therefore, the Ideal Observer judges the center to be far for all values of flicker f. Note also that the expression is independent of *f*. This means that the confidence of the Ideal Observer is unaffected by flicker.

Section S3. Perception of AD stimuli as a function of eccentricity



Supplemental Figure S3. Depth-order perception of our AD stimuli as a function of stimulus eccentricity. (**a,b**) DFM percepts elicited by the same pair of stimuli when they are viewed foveally (**a**) and parafoveally at an eccentricity of 5° (**b**).

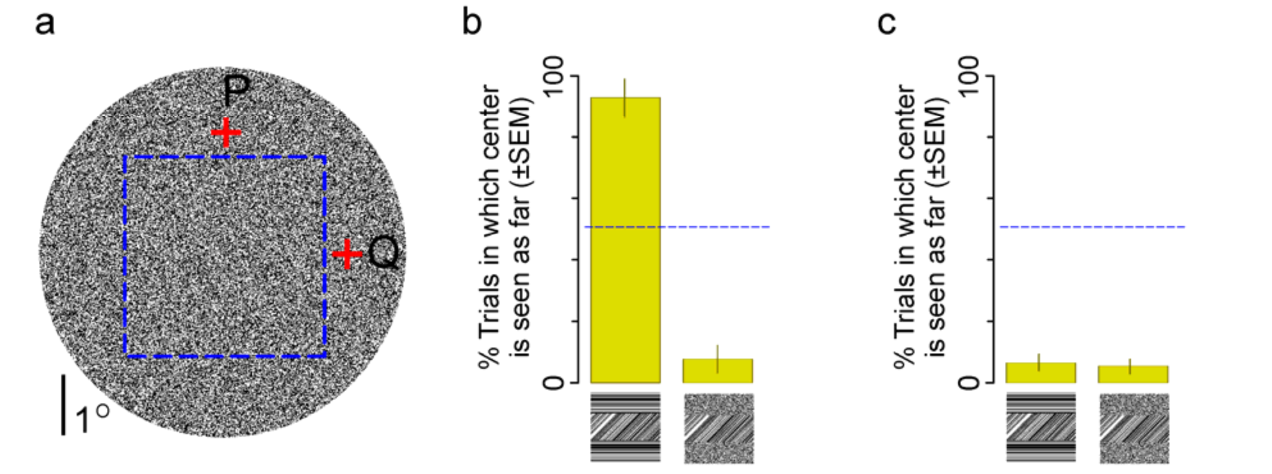
Our preliminary results (not shown) indicated that when the stimuli are viewed parafoveally, depth-order is perceived incorrectly even more often than with foveal viewing. Here we document this effect quantitatively and argue that this does not weaken the arguments presented in the main text.

Panel a in Fig. S3 replicates the effect shown in Fig. 2 of the main text, in which the stimuli were viewed foveally. As expected, the depth-order reversed when the surround is static *vs*. flickery (Fig. S3a left *vs*. right, respectively). However, when the same pair of stimuli is viewed parafoveally, the depth-order does not reverse for flickery surrounds, *i.e.,* the center is perceived as near regardless of whether the surround is stationary or flickery (Fig. S3b left and right bars, respectively).

It is important to emphasize that this lack of depth-order reversal does not in any way diminish or disprove our arguments regarding the insufficiency of the AD cue. Note that with parafoveal viewing, *both* stimuli elicit percepts that are *opposite* of the percept predicted by the AD cue. Thus, if anything, this strengthens our arguments about the insufficiency of the AD cue by presenting another case in which the AD cue fails to dictate the DFM percept.

It is noteworthy that the stimuli in Fig. S3b do not elicit chance-level performance, but consistently elicit near percepts instead. We show below (Section S4) evidence that indicates that a possible explanation for this effect is that the visual system has a perceptual bias to interpret shear as a nearness cue.

Section S4. Biasing effects of motion shear



Supplemental Figure S4. Depth-order perception of our AD stimuli as a function of the position of fixation relative to the AD border *vs*. shear border. See text for details.

In this experiment, the center was a square. When the center dots move vertically (similar results were obtained for other directions of movement), the top and bottom sides of the center represent the AD border, or the border at which the dots undergo accretion/deletion. Similarly, the vertical sides represent the shear border, along which the center dots are in a shear motion relative to the surround dots.

When the subjects fixated along the horizontal border (fixation position P), the results (Fig. S4b) essentially replicated those in Fig. 2 of the main text. Note, incidentally, that this also indicates that the square stimulus used in the Ideal Observer analysis can elicit the relevant results.

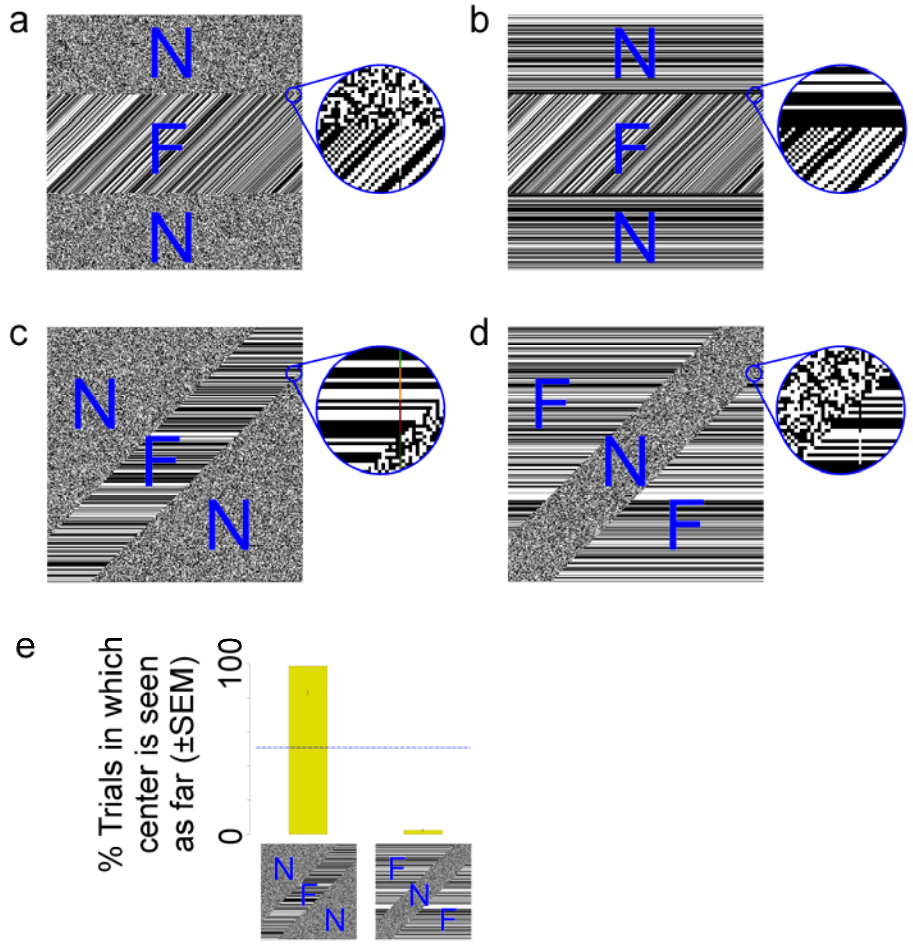
When the subjects fixated along the vertical border (fixation position Q), which also had the effect of increasing the eccentricity of the AD border, the subjects reliably reported near percepts (Fig. S4c).

The above effects can also be elicited by fixating at the center of the stimulus and attending to horizontal *vs*. vertical borders (data not shown). (Also see Demo Movie 2.) These results are consistent with the results of Royden and colleagues who have shown similar effects when the shear border is made more prominent than the AD border [10].

It is important to emphasize that shear is solely a perceptual bias, and is not a stimulus-based DFM cue like the AD and BF cue are. This is because the particular type of shear that elicits the depth-order bias in this case, in which one surface is stationary while the other is moving, can just as easily result when the moving surface is nearer or farther than the stationary surface.

Taken together, these considerations suggest that when the stimulus-driven DFM evidence (in this case, the AD cue) is weak, instead of interpreting the stimulus randomly (*i.e.*, at chance levels), the brain relies more on its perceptual biases based on the shear cue. Thus, the DFM percept at any given viewing location represents the relative balance between the strengths of stimulus-driven, or 'bottom-up' cues *vs*. 'top-down' signals such as the near bias resulting from shearing motion.

Section S5. Flickering occluder does not necessarily diminish the AD cue, and a flickering surface is not necessarily reported as the far surface.



Supplemental Figure S5. AD cue with or without the motion of the AD border. See text for details.

As noted in the main text, one potential concern about our depth-order illusion is that introducing the flicker in the surround may somehow affect the accretion-deletion information in the center. Using conventional optical flow techniques, we have shown in Fig. 3 of the main text that the flicker in the surround does not affect the available accretion-deletion information in the center in our stimuli. Here we provide psychophysical evidence to further support this view.

Figure S5 (a) and (b) are ST plots of the two main stimuli used in our study. Both represent a static occluding surface and a moving occluded surface. The corresponding insets show that the termination of motion trajectories, which constitute the AD cue ([1,2]; also see Fig. S1 above), are unaffected by the flicker.

Previous studies have shown that when the occluder moves, the AD cue can elicit the predicted percepts, *even when the occluder is flickery* [1,2]. Figure S5 (c) and (d) both show such stimuli with flickering, but moving, occluders. The sole DFM in these stimuli is the AD cue (see Fig. 2 of ref. [1,2]).

Note that the motion trajectory terminators in these stimuli are directly comparable to those in panels a and b (insets). Thus, if flickering occluder affects the AD cue in the occluded object, then the actual percepts should be different than those predicted by the AD cue (blue letters; N=near, F=far). Subjects reported the depth-order of the middle surface (which represents the occluded surface in Fig. S5c and the occluder in Fig. S5d). However, the reported percepts were entirely consistent with the predicted percepts. Taken together, these results indicate that the flickering occluder by itself does not diminish the perceptual efficacy of the AD information. In other words, the motion terminators do remain interpretable as AD cue. Rather, what diminishes the AD cue when the occluder is static in our stimuli is the absence of strong cues about the AD border. When the surround is static and flickery, the border is not clearly distinguishable (which can be ascertained from Demo Movie 2). However, when the occluder moves, even though it remains flickery, it clearly delineates the border between the occluder and the occluded surface. Thus, information unrelated to accretion-deletion *per se*, namely segmentation information that delineates the occluded surface from the occluder, is needed to constrain the AD cue. This also helps explain when the border between the two surfaces is delineated by static segmentation cues such as contrast, color and luminance, it restores the DFM percepts predicted by the AD percept (see main text).

The results shown in Fig. S5 also address a potential concern about the Moonwalk illusion. Recall that in the illusion, the perceived nearness of the center is closely linked with the flickering surround, in that the subjects reported the center as near when the surround was flickering, and center as far when the surround was not flickering. One possible confounding explanation for this effect is that the subjects spuriously associated the surround flicker (or the presence of flicker anywhere in the stimulus) with the nearness of the center, and reported their depth-order percepts accordingly. This confound is highly unlikely, for three reasons. First, the depth-order reports of the subjects are self-evidently valid, in that the stimuli do elicit the corresponding reported percepts, as can be verified from Demo Movies 1 and 2. Second, subjects had no reason to form the aforementioned spurious associations, since they were told to report what they saw, and that there was no 'correct' percept *per se*. Thus, for the above confound to be valid, all subjects would have had to ignore the instructions and formed the exact same spurious association. Third, as shown in Fig. S5, the same subjects also reported the flickering surface as the near surface in other stimuli that contained a flickering surface, indicating that the subjects did not necessarily report flickering surfaces as far, or report the center as near whenever there was flicker anywhere in the stimulus. Taken together, these considerations indicate that the Moonwalk percept is not attributable to this confound.

SUPPLEMENTAL REFERENCES

1. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2: 284-299.
2. Hegdé J, Albright TD, Stoner GR (2004) Second-order motion conveys depth-order information. J Vis 4: 838-842.
3. Weiss Y, Simoncelli EP, Edelson EH (2002) Motion illusions as optimal percepts. Nat Neurosci 5: 598-604.
4. Wallach H (1948) Brightness constancy and the nature of achromatic colors. J Exp Psychol 38: 310-324.
5. Leibowitz H, Myers NA, Chinetti P (1955) The role of simultaneous contrast in brightness constancy. J Exp Psychol 50: 15-18.
6. Ullman S (1979) The interpretation of Visual Motion. Cambridge, MA: MIT Press.
7. Gibson JJ, Kaplan GA, Reynolds HEN, Wheeler K (1969) The change from visible to invisible: A study of optical transitions. Percept Psychophys 5: 113-116.
8. Kaplan GA (1969) Kinetic disruption of optical texture: the perception of depth at an edge. Perception and Psychophysics 6: 193-198.
9. Howard IP, Rogers BJ (2002) Seeing in Depth. Vol. 2. Depth Perception: I. Porteous, Toronto.
10. Royden CS, Baker JF, Allman J (1988) Perceptions of depth elicited by occluded and shearing motions of random dots. Perception 17: 289-296.